UNU-MERIT Working Paper Series

#2014-035

Stability and strategic diffusion in networks

Théophile T. Azomahou and Daniel C. Opolot

Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)
email: info@merit.unu.edu | website: http://www.merit.unu.edu

Maastricht Graduate School of Governance (MGSoG)
email: info-governance@maastrichtuniversity.nl | website: http://mgsog.merit.unu.edu

Keizer Karelplein 19, 6211 TC Maastricht, The Netherlands
Tel: (31) (43) 388 4400, Fax: (31) (43) 388 4499

UNU-MERIT Working Papers

ISSN 1871-9872

Maastricht Economic and social Research Institute on Innovation and Technology, UNU-MERIT

Maastricht Graduate School of Governance MGSoG

UNU-MERIT Working Papers intend to disseminate preliminary results of research carried out at UNU-MERIT and MGSoG to stimulate discussion on the issues raised.

Stability and strategic diffusion in networks

Théophile T. Azomahou

Maastricht University—UNU-MERIT, 6211 TC Maastricht, The Netherlands

Daniel C. Opolot*

Maastricht University—UNU-MERIT, 6211 TC Maastricht, The Netherlands

This Version—May 2014

Abstract: Learning and stochastic evolutionary models provide a useful framework for analyzing repeated interactions and experimentation among economic agents over time. They also provide sharp predictions about equilibrium selection when multiplicity exists. This paper defines three convergence measures, *diffusion rate, expected waiting time* and *convergence rate*, for characterizing the short-run, medium-run and long-run behavior of a typical model of stochastic evolution. We provide tighter bounds for each without making restrictive assumptions on the model and amount of noise as well as interaction structure. We demonstrate how they can be employed to characterize evolutionary dynamics for coordination games and strategic diffusion in networks. Application of our results to strategic diffusion gives insights on the role played by the network topology. For example we show how networks made up of cohesive subgroups speed up evolution between quasi-stable states while sparsely connected networks have the opposite effect of favoring almost global stability.

Keywords: Learning and evolution, networks, diffusion rate, convergence rate, expected waiting time

JEL classification: C73, D80

1. Introduction

Learning and evolutionary models provide a framework for modeling repeated interactions and experimentation among economic agents over time. They also provide a useful framework for identifying the most likely long-run outcomes in economic environments with multiple possible outcomes.[1] The basic idea of learning and evolutionary dynamics is that agents learn their opponents' play and subsequently adjust their behavior over time. The adjustment process is such that each agent plays their optimal strategy with higher probability and with a small probability randomizes among the elements of the strategy set. The analysis then usually focuses on establishing the long-run behavior

*Corresponding author: Opolot D. C., Maastricht University—UNU-MERIT, Keizer Karelplein 19, 6211 TC Maastricht, The Netherlands, (opolot@merit.unu.edu), Tel. +31 433884440, Fax +31 433884499.

[1]Several applications of learning and evolutionary game theory that have been developed include bargaining, signaling, contagion and social innovation (e.g. Vega-Redondo (1997), Nöldeke and Samuelson (1997), Kandori and Rob (1998), Alós-Ferrer and Weidenholzer (2008), Huck et al. (2012)).

of the learning process, in which solution concepts such as *stochastic stability* are employed (Foster and Young, 1990; Kandori et al., 1993; Young, 1993; Ellison, 2000).

In the computation of stochastically stable outcomes (sets) it is customary to assume vanishing probability of mistakes occurring. But as pointed out by Ellison (1993) and Sandholm (2010), this assumption also implies that the convergence time to the long-run stationary distribution becomes exceptionally long. More specifically, once the process gets trapped in one of the quasi-stable states (possibly a socially undesirable one), the waiting time to exit its *basin of attraction* becomes arbitrarily large for vanishing noise. If on the other hand the level of noise is allowed to be positive, then the most stable outcomes are not necessary those that are predicted by stochastic stability. A typical example of this case is when the probabilities of mistakes are state-dependent (Bergin and Lipman, 1996).

The present paper aims to provide general convergence measures that can be employed to characterize the short-run, medium-run and long-run behavior of a typical model of learning and evolutionary game dynamics. More specifically, we define and provide bounds for *diffusion rate*, *expected waiting time* and *convergence rate* as measures for short-run, medium-run and long-run behavior respectively. The diffusion rate measures how fast or slow a given behavior, action or strategy gets adopted by a subset or entire population once its minimum threshold has been reached. The expected waiting time measures how long it takes the learning process to exit the boundary of the basin of attraction of a given (quasi-) stable state or subset of states. The diffusion rate and expected waiting time can thus together be used to characterize evolution between two or more stable states. The convergence rate on the other hand measures how fast the process attains its long-run invariant distribution.

Unlike convergence measures in the literature, for example Ellison (1993), Binmore and Samuelson (1997), Montanari and Saberi (2010) and those generally employed in the computation of stochastic stability (e.g. Young (1993), Kandori et al. (1993), Ellison (2000), and Staudigl (2012) among others), the bounds we provide for diffusion rate, expected waiting time and convergence rate do not rely on the assumption of vanishing noise. The generality of our framework also allows for heterogeneity in the probabilities of mistakes across the population; for example it can be state-dependent or depend on each agent's position in the network in the case of local interactions. We specifically derive bounds for diffusion rate in terms of the spectral properties of the interaction structure and the parameters of the underlying game. The bound for the expected waiting time of any given quasi-stable set is a function of the size of its basin of attraction and aggregate (average) probability of the mistakes actually occurring. The direct implication is that the most stable set is not necessarily that with the largest basin of attraction, which is contrary to the findings of the case in which assumption of vanishing noise is made.

We demonstrate how these three convergence measures can be employed to characterize the properties of an evolutionary process in binary coordination games and strategic diffusion in networks. There has been a growing body of literature on evolutionary dynamics and strategic diffusion in networks for example that on deterministic best response dynamics (Morris, 2000; Lee et al., 2003), noisy best response dynamics (Ellison, 1993; Blume, 1995; Alós-Ferrer and Weidenholzer, 2008; Young, 2011; Montanari and Saberi, 2010) and mean field best response dynamics (Sandholm, 2001; Alós-Ferrer and Weidenholzer, 2008; Lelarge, 2012). Application of the results for

the three convergence measures provides complementary findings to those in the above papers. For example we find that the expected waiting time under global interactions grows exponentially with population size while under local interactions it is independent of population size. But in addition to this well established result, we are able to demonstrate how the presence of cohesive subgroups in the network speeds up evolution between quasi-stable states (sets). This effect is different from that of local interactions mentioned above. In fact, we find that sparsely connected networks reinforce almost global stability and reduce the rate of inter-quasi-stable states transitions. For example in binary coordination games, sparsely connected networks ensure faster evolution toward the risk-dominant choice and reduce the likelihood of exiting from it. A combination of our results on expected waiting time and diffusion rates also implies that strategic diffusion is generally faster in networks consisting of cohesive subgroups. These cohesive subgroups should be small enough to favor shorter expected waiting time but large enough and sufficiently connected to favor higher diffusion rate within them.

The remainder of the paper is organized as follows. In section 2 we introduce the general framework of stochastic evolutionary dynamics in general interaction structures. Section 3 is devoted to the characterization of expected waiting time. Section 4 applies the results of Section 3 to binary pure coordination games and Section 5 focuses on analysis of diffusion and convergence rates. All proofs are relegated to the Appendix.

2. The model

We consider a general class of $n-$person repeated games that exhibit strategic complementarity. A set of agents $N = \{1, \cdots, i, \cdots, n\}$ interact through a social network, which includes a possibility of global interactions. We define the network of interactions in a graph theoretic manner. Let $G(n, E)$ be a graph with n vertices representing the number of agents and E edges linking different pairs of agents, such that a graph g_{ij} defines the connection between i and j. If $g_{ij} = 1$ then a directed link exists from i to j, and zero implies otherwise. $G(n, E)$ is thus a directed network describing the relationship of any one agent with every other agent in the population. The *adjacency matrix* $\mathscr{A}$ of $G(n, E)$ is defined as an $n \times n$ matrix with entries being the elements of g_{ij}. The *neighborhood* of agent i, $\mathscr{N}_i$, is defined as $\mathscr{N}_i = \{j \in N | g_{ij} = 1\}$, and gives the set of players to which i is linked to. The cardinality $\#\mathscr{N}_i = k_i$, is the *degree* of i.

The set of actions X_i for each $i \in N$ is finite. We denote the action chosen by each $i \in N$ at period t by $x_{i,t}$, and for all $x_{i,t} \in X_i$, $\mathbf{x}_t = (x_{1,t}, \cdots, x_{n,t})$ denotes the strategy profile at t. Each $\mathbf{x}_t \in \mathbf{X} = \prod_{i=1}^{n} X_i$ will also be referred to as the *population state* or simply the *state* of the learning process at t, where $\mathbf{X}$ is the state space.

2.1. Payoff structure

For each $i \in N$, the functions $u_i : \mathbf{X} \to \mathbb{R}$ define the payoffs of the game, such that $u_i(x_i, \mathbf{x}_{-i})$ is the payoff of i when he plays x_i and the other players follow strategy profile $\mathbf{x}_{-i}$. Games of strategic complements that we consider include pure coordination games such as that in Table 1 below, and those with additively separable payoff functions. In the case of pure coordination games,

let $v_i(x_i, x_j)$ be the payoff to i from playing action x_i when his opponent $j \in \mathcal{N}_i$ plays actions x_j. Then the total payoff to i from playing x_i when the other players follow strategy $\mathbf{x}_{-i}$ is of the form.

$$u_i(x_i, \mathbf{x}_{-i}) = \sum_{j \in \mathcal{N}_i} J_{ij} v_i(x_i, x_j), \tag{1}$$

The parameter J_{ij} depends on whether players are randomly and uniformly matched with every

Table 1: Payoff structure for the pure coordination game between i and j

		player j: A	player j: B
player i: A		a , a	d , c
player i: B		c , d	b , b

other in the population (global interactions) in which case $J_{ij} = \frac{1}{n}$, or they interact locally through a network. Note that local interactions can also be random and uniform, in which case $J_{ij} = \frac{1}{k_i}$. Otherwise, $J_{ij} \in [0, 1]$ for all $j \in \mathcal{N}_i$ and for each $i \in N$.

The case of additively separable utilities is generally represented by linear-quadratic functions of the form

$$u_i(x_i, \mathbf{x}_{-i}) = s_i(x_i) + \sum_{j \in \mathcal{N}_i} S_i^j(x_i, x_j). \tag{2}$$

where $u_i(x_i)$ is the intrinsic utility to i from playing strategy x_i and $S_i^j(x_i, x_j)$ is the network externality or social utility to i from playing action x_i when the neighbor j's action is x_j. The notable examples include the *status* model of Akerlof (1997), the social interactions model of Brock and Durlauf (2001) and the neighborhood effects models discussed in Glaeser and Scheinkman (2001).

2.2. Revision probabilities

Agents revise their actions following Darwinian dynamics. That is they respond myopically to the past strategies of their opponents, playing the "optimal" strategy with high probability and with a small probability they play that which is not necessarily optimal. By "optimal" we mean a strategy that would be prescribed by a given learning rule. For example under Best-Reply dynamics it would be the strategy which maximizes the associated utility function, and under imitation dynamics it would be that which is the most successful in the population or neighborhood. We focus on the case of Best-Reply dynamics in this paper.

The main results we derive generally apply to the dynamics with revision probabilities of the form

$$\mathbb{P}_i(x_{t+1} = x | \mathbf{x}_t) = (1 - \varepsilon_i(\mathbf{x}_t)) BR_i(x_{t+1} = x | \mathbf{x}_t) + \varepsilon_i(\mathbf{x}_t) \mathcal{P}_i(x_{t+1} = x | \mathbf{x}_t) \tag{3}$$

where $BR_i(x_{t+1} = x|\mathbf{x}_t)$ is the probability that i plays action x in the next period under best-reply dynamics given that the current state is $\mathbf{x}_t$. That is

$$BR_i(x_{t+1} = x|\mathbf{x}_t) = \begin{cases} 1 & \text{if } x \in \arg\max_{x_i \in X_i} u_i(x_i, \mathbf{x}_t) \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

For each $i \in N$, $\varepsilon_i(\mathbf{x})$ is the state-dependent mutation rate, the probability that i randomizes among the elements of X_i with the conditional distribution defined by $\mathcal{P}_i(x|\mathbf{x})$. The product $\mathscr{P}_i(x|\mathbf{x}_t) = \varepsilon_i(\mathbf{x}_t)\mathcal{P}_i(x_{t+1} = x|\mathbf{x}_t)$ is the probability of playing action x by mistake. We denote the vector of mutation rates by $\boldsymbol{\varepsilon} = (\varepsilon_1(\mathbf{x}), \cdots, \varepsilon_n(\mathbf{x}))$. Note that for each $i \in N$ and $\mathbf{x}_t \in \mathbf{X}$, $\sum_{x \in X_i} \mathcal{P}_i(x_{t+1} = x|\mathbf{x}_t) = 1$.

In the application to strategic diffusion that will be provided in the sections that follow, we shall focus on the *logit* dynamic rule, which is of the form

$$\mathbb{P}_i(x_{t+1} = x|\mathbf{x}_t) = \frac{\exp\left[\beta u_i(x, \mathbf{x}_t)\right]}{\sum_{y \in X_i} \exp\left[\beta u_i(y, \mathbf{x}_t)\right]} \tag{5}$$

where $\beta > 0$ is a parameter of randomness in agents' decision process. The smaller β the more prone agents are to mistakes. Unlike other evolutionary game models in which the probabilities of mistakes are uniform (e.g. Kandori et al. (1993), Young (1993) and Ellison (1993)), logit dynamics exhibits state-dependent mutation rates. More specifically, the probability of making an error is a function of the payoff loss associated with that error. For example in the case of binary strategy set $X = \{A, B\}$

$$\mathbb{P}(x_{i,t+1} = A|\mathbf{x}_t) = \frac{\exp\left[\beta v_i(A, \mathbf{x}_t)\right]}{\exp\left[\beta v_i(A, \mathbf{x}_t)\right] + \exp\left[\beta v_i(B, \mathbf{x}_t)\right]} = \frac{1}{1 + \exp[\beta \Delta v_i]}. \tag{6}$$

where $\Delta v_i = v_i(B, \mathbf{x}_t) - v_i(A, \mathbf{x}_t)$ is the payoff loss from playing strategy A rather than B. Logit revision rule also allows for the possibility of probabilities of mistakes to depend on an agent's position in the network. Consider an example of a binary pure coordination game in Table 1 in which $a = 2, b = 1, c = d = 0$. If the interactions are such that agents are randomly and uniformly matched (that is $J = \frac{1}{n}$ or $\frac{1}{k_i}$) then the probability of making a mistake to play A when all agents are initially playing B is

$$\mathbb{P}_i(x = A|\vec{B}) = \frac{1}{1 + \exp(\beta)} \quad \text{for all } i \in N. \tag{7}$$

where $\vec{B}$ implies all B population state. If on the other hand each agent's payoff is the sum of neighbors' choices (that is $J = 1$) then

$$\mathbb{P}_i(x = A|\vec{B}) = \frac{1}{1 + \exp(\beta k_i)}. \tag{8}$$

In which case highly connected agents have a lower probability of making an error than less connected ones.

The evolution of the population state for a given interaction structure can thus be treated as a Markov chain $(\mathbf{X}, P_\varepsilon)$, defined on the finite set of states $\mathbf{X}$ with a fixed stochastic transition matrix P_ε whose elements are defined by

$$P_\varepsilon(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{n} \mathbb{P}_i(x_{t+1} = y_i|\mathbf{x}_t = \mathbf{x}) \quad \text{for each } y_i \in \mathbf{y} \tag{9}$$

2.3. *Recurrent classes and basins of attraction*

The perturbed process $(\mathbf{X}, P_\varepsilon)$ has a unique invariant (stationary) distribution $\pi_\varepsilon = \lim_{t\to\infty} \mathbf{q}_0 P_\varepsilon^t$, where $\mathbf{q}_t$ is the vector of probability mass functions at period t . The existence of a unique stationary distribution is a standard fact about *aperiodic-irreducible* Markov chains. The stationary distribution describes the amount of time the process spends in each state in the long-run. The *communication recurrent classes* of $(\mathbf{X}, P_\varepsilon)$ are the *limit sets* of the identical process $(\mathbf{X}, P)$ without mistakes.[2] In the coordination game of Table 1 above for example, the limit sets include the singleton sets in which all players play strategy A and in which they all play strategy B. Generally, the limit sets of $(\mathbf{X}, P)$ can include sets that are cycles and those in which players use different strategies. When the interactions are governed by a social network , the number of limit sets is enhanced. In particular, there will exists singleton limit sets in which strategies co-exists and different *cohesive* subgroups adopt different strategies. The limit set of $(\mathbf{X}, P)$ that results depends on the initial state of the process. We denote the typical limit set of $(\mathbf{X}, P)$ and hence recurrent class of $(\mathbf{X}, P_\varepsilon)$ by Ω_j, and by $\mathbf{\Omega}$ for the set of all possible Ω_j.

The *basin of attraction* $D(\Omega)$ of a set Ω is defined as $D(\Omega) = \{\mathbf{x} \in \mathbf{X} | \mathbb{P}(\exists T \text{ s.t } \mathbf{x}_t \in \Omega \ \forall\ t > T | \mathbf{x}_0 = \mathbf{x}) = 1\}$. Without loss of generality, the model of mistakes is such that the basins of attraction of the recurrent sets of $(\mathbf{X}, P_\varepsilon)$ are equivalent in composition to those of the unperturbed process $(\mathbf{X}, P)$. Basins of attraction induce a partition on the state space into disjoint subsets $D(\Omega) \subset \mathbf{X}$. Let $\tilde{\mathbf{x}}$ be the shorthand for $D(\Omega)$ and let $\tilde{\mathbf{X}}$ be the state space consisting of $\tilde{\mathbf{x}}$'s as its states. The behavior of the process $(\mathbf{X}, P_\varepsilon)$ can thus be fully characterized by establishing the measures of persistence and attractiveness of its recurrent classes and hence basins of attraction. In the next section we focus on the measures related to persistence of recurrent classes.

3. Resistance and Expected waiting time

In this section, we define *resistance* and *Expected waiting time* as measures associated with the persistence of recurrent classes. There are two factors that determine the persistence of a recurrent class, the distance to the neighboring basins of attraction—*diameter*, and the actual aggregate probability of mistakes occurring. Define the (normalized) diameter $d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)$ of the directed relation $\tilde{\mathbf{x}}_i \to \tilde{\mathbf{x}}_j$ as the fraction of mistakes required to enter the basin of attraction of $\tilde{\mathbf{x}}_j$ starting from the recurrent class Ω_i of $\tilde{\mathbf{x}}_i$. Equivalently, $d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)$ is the fraction of players required to simultaneously play a different action by mistake for the process $(\mathbf{X}, P_\varepsilon)$ to enter the boundary of $\tilde{\mathbf{x}}_j$ given that it is in the state $\mathbf{x} \in \Omega_i$.

The collective probability associated with the diameter $d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)$ or a direct transition $\tilde{\mathbf{x}}_i \to \tilde{\mathbf{x}}_j$ is captured by its average probability. Formally, let $\mathbb{P}_i(x_{t+1} = y | \mathbf{x}_t = \mathbf{x})$ for $y \in \mathbf{y} \in \Omega_j \in \tilde{\mathbf{x}}_j$ and $y \notin \mathbf{x} \in \Omega_i \in \tilde{\mathbf{x}}_i$ be the probability that i plays an action belonging to the state in a recurrent class $\Omega_j \in \tilde{\mathbf{x}}_j$ that is different from that played in the state belonging to the recurrent class $\Omega_i \in \tilde{\mathbf{x}}_i$,

[2] A set $\Omega \subset \mathbf{X}$ is a limit set of $(\mathbf{X}, P)$ if $\forall \mathbf{y} \in \Omega$, $\mathbb{P}(\mathbf{x}_{t+1} \in \Omega | \mathbf{x}_t = \mathbf{y}) = 1$, and that $\forall \mathbf{y}, \mathbf{z} \in \Omega$, there exists a $\tau > 0$ such that $\mathbb{P}(\mathbf{x}_{t+\tau} = \mathbf{z} | \mathbf{x}_t = \mathbf{y}) > 0$

given that $(\mathbf{X}, P_\varepsilon)$ is in $\mathbf{x} \in \Omega_i$. We can the define the average probability $\mathbb{P}_A(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)$ as

$$\mathbb{P}_A(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}_i(x_{t+1} = y | \mathbf{x}_t = \mathbf{x}) \quad \text{for } y \in \mathbf{y} \in \Omega_j \text{ and } y \notin \mathbf{x} \in \Omega_i \tag{10}$$

Given the diameter and the average probability, we then define the *cost* $c(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)$ of a direct transition $\tilde{\mathbf{x}}_i \to \tilde{\mathbf{x}}_j$ as $c(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) = -d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) \ln(\mathbb{P}_A(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j))$. The resistance $\mathscr{R}(\tilde{\mathbf{x}}_i)$ is defined as the minimum cost over all $\tilde{\mathbf{x}}_j \neq \tilde{\mathbf{x}}_i$. That is

$$\mathscr{R}(\tilde{\mathbf{x}}_i) = \min_{\tilde{\mathbf{x}}_j \neq \tilde{\mathbf{x}}_i} \{c(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)\}$$

Expected waiting time is a measure of how long it takes the process $(\mathbf{X}, P_\varepsilon)$ to exit the basin of attraction of a recurrent class once it has entered its boundaries. Equivalently, it is the measure of the persistence of the limit sets of $(\mathbf{X}, P)$ to perturbations and can thus be used to rank limit sets in terms of how locally stable they are. Formally,

Definition 1: *Let Ω_i and Ω_j be two recurrent classes of the process $(\mathbf{X}, P_\varepsilon)$ with respective basins of attraction $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{x}}_j$. Let $n(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)$ be the number of agents that must play a different action for the transition from Ω_i to the boundary of $\tilde{\mathbf{x}}_j$ to occur. Then the Expected waiting time $\mathscr{T}(\Omega_i)$ of Ω_i associated with such a transition is defined as*

$$\mathscr{T}(\Omega_i) = \min_{\tilde{\mathbf{x}}_j \neq \tilde{\mathbf{x}}_i} \mathbb{E}\left[\min\{t |\ n(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) \geq d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)n\}\right]$$

The following theorem provides a lower bound on Expected waiting time.

Theorem 1: *Let $(\mathbf{X}, P_\varepsilon)$ be a model of learning with mistakes, and let Ω_i and Ω_j be any two of its recurrent classes with corresponding basins of attraction $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{x}}_j$. Then*

$$\mathscr{T}(\Omega_i) \geq e^{n[f(d(\tilde{\mathbf{x}}_i)) + \mathscr{R}(\tilde{\mathbf{x}}_i)]}.$$

where $f(a) = a \ln\left(\frac{a}{1-a}\right) + \ln(1-a)$, and $d(\tilde{\mathbf{x}}_i) = \arg\min_{d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)} c(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)$

Proof. See Appendix A.1 □

The expected waiting time of a given recurrent class is an exponential function of the resistance of its basin of attraction and the diameter $d(\tilde{\mathbf{x}}_i)$ that minimizes the cost of exiting its basin of attraction. Note that the size of basin of attraction is the minimum number (or fraction) of mistakes required to exit its boundaries. That is for any $\tilde{\mathbf{x}}_i \in \tilde{\mathbf{X}}$, the size of basin of attraction of $\tilde{\mathbf{x}}_i$ is given by $\min_{\tilde{\mathbf{x}}_j \neq \tilde{\mathbf{x}}_i} d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)n$. The direct implication of Theorem 1 is that depending on the model of mistakes, the recurrent class with the largest size of basin of attraction is not necessarily that with the highest expected waiting time. This finding is contrary to that derived by assuming vanishing noise such as Ellison (1993), in which the expected waiting time depends solely on the size of its basin of attraction. In the next section we demonstrate how the result in Theorem 1 can be used to characterize the behavior of an evolutionary process for binary pure coordination games.

4. Pure coordination games

This section applies the results of Theorem 1 to characterize the behavior of evolutionary game dynamics in pure coordination games. We focus on the binary action case with payoff of the form in Table 1, and that in which agents are randomly and uniformly matched (both global $J_{ij} = \frac{1}{n}$ and local interactions $J_{ij} = \frac{1}{k_i}$).

4.1. *Global interactions*

Consider the coordination game in Table 2 played by a sufficiently large number of players. Under global interactions, there exists only two recurrent communication classes $\vec{A}$ and $\vec{B}$ with respective basins of attraction $\tilde{A}$ and $\tilde{B}$. When agents interact randomly and uniformly, the transition probabilities are identical and independent of their positions in the network. That is

$$\mathbb{P}_i(x_{t+1} = B | \mathbf{x}_t = \vec{A}) = \frac{1}{1 + \exp(\beta(1+\alpha))} \tag{11}$$

for all $i \in N$. Implying that $\mathbb{P}_A(\tilde{A}, \tilde{B}) = 1/(1+\exp(\beta(1+\alpha)))$. Similarly, $\mathbb{P}_A(\tilde{B}, \tilde{A}) = 1/(1+\exp(\beta))$.

Table 2: For $\alpha > 0$, A is both payoff and risk dominant action.

	A	B
A	$1+\alpha$, $1+\alpha$	0 , 0
B	0 , 0	1 , 1

The corresponding normalized diameters are $d(\tilde{A}, \tilde{B}) = \frac{1+\alpha}{2+\alpha}$ and $d(\tilde{B}, \tilde{A}) = \frac{1}{2+\alpha}$. That is starting from $\vec{A}$, at least a fraction $\frac{1+\alpha}{2+\alpha}$ of the population must first switch to playing B for a global cascade to $\vec{B}$ to occur. Similarly $\frac{1}{2+\alpha}$ must first adopt action A for a global cascade to $\vec{A}$ to occur. The corresponding resistances are thus

$$\mathscr{R}(\tilde{A}) = \frac{1+\alpha}{2+\alpha} \ln\left(1 + e^{\beta(1+\alpha)}\right)$$

$$\mathscr{R}(\tilde{B}) = \frac{1}{2+\alpha} \ln\left(1 + e^{\beta}\right)$$

Consequently, the expected waiting time to exit each basin of attraction are given by

$$\mathscr{T}(\vec{A}) \geq e^{n\left[\frac{1+\alpha}{2+\alpha}\ln\left(1+e^{\beta(1+\alpha)}\right)+f\left(\frac{1+\alpha}{2+\alpha}\right)\right]} > e^{n\left[\frac{1}{2+\alpha}\ln\left(1+e^{\beta}\right)+f\left(\frac{1}{2+\alpha}\right)\right]} \leq \mathscr{T}(\vec{B})$$

Implying that under global interactions the expected waiting time is an exponential function of the population size. It also implies that for a finite population, the most likely outcome for logit dynamics under global interactions is that in which all agents play the risk-dominant action. This however is not necessarily true for other learning rules or multi-action coordination games.

4.2. *Local interactions: Deterministic networks*

The results in Theorem 1 can be employed to gain a richer understanding on how the network topology affects the process of evolutionary game dynamics. The network topology affects both the diameters and average probabilities associated with inter-basin of attraction transitions. For the case of deterministic and/or simple network topologies, the diameter and average probabilities can be computed by considering the contributions of each player relative to their degree. To begin with,

consider n agents arranged in a circle such that each interacts with two others on the left and right. If agents play the pure coordination game in Table 2 and that they are randomly and uniformly matched with their neighbors, then the average probabilities are as above. For $\alpha > 0$, we have that $d(\tilde{A}, \tilde{B}) = 1$ and $d(\tilde{B}, \tilde{A}) = \frac{1}{n}$. That is, starting from $\vec{A}$ each agent requires both neighbors to switch to B before doing so. Hence, to evolve from $\vec{A}$ to $\vec{B}$ all agents must simultaneously switch to B. On the other hand, evolving from from $\vec{B}$ to $\vec{A}$ requires only a single mutation. The corresponding expected waiting times then become.

$$\mathscr{T}(\vec{A}) \geq e^{n \ln\left(1+e^{\beta(1+\alpha)}\right)} > Ce^{\ln\left(1+e^{\beta}\right)} \leq \mathscr{T}(\vec{B})$$

where $C = e^{nf(\frac{1}{n})}$. Clearly, as n increases, the only most likely outcome is $\vec{A}$, in which all agents play the payoff and risk dominant action. Compared to the global interactions case above, the expected waiting time to exit $\tilde{B}$ is independent of the population size. A result that has also been established by other authors such as Ellison (1993) and Lee et al. (2003).

Now consider a set of eight agents whose interactions are governed by the network topology in Figure 1, and playing a pure coordination game in Table 2. The network can be divided into two cohesive subgroups $g_1 = \{1, 2, 3, 4\}$ and $g_2 = \{5, 6, 7, 8\}$. A player i belongs to a given cohesive subgroup if and only if at least half of i's interactions are with members of that subgroup. In addition to $\vec{A}$ and $\vec{B}$, there are two additional recurrent classes; $\vec{A}_1$ and $\vec{A}_2$ in which all members of g_1 play A while those in g_2 all play B and vise versa for $\vec{A}_2$. The corresponding normalized diameters

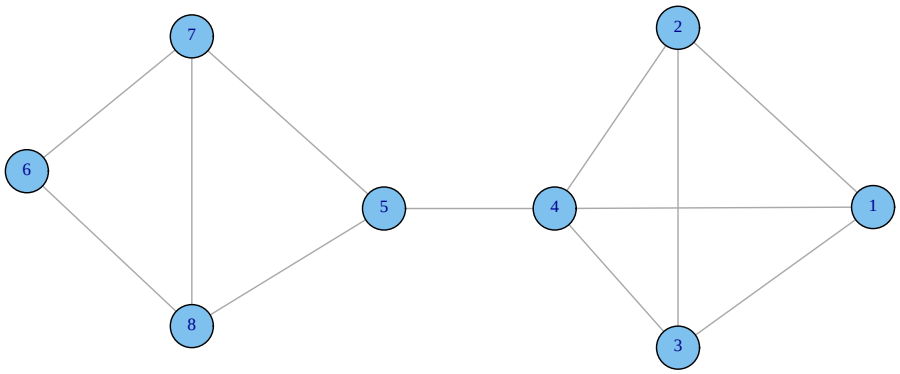

Figure 1: A general network with two cohesive subgroups

are then given by $d(\tilde{B}, \tilde{A}_1) = \frac{1}{4}$, $d(\tilde{B}, \tilde{A}_2) = \frac{1}{8}$, $d(\tilde{A}, \tilde{A}_1) = \frac{3}{8}$, $d(\tilde{A}, \tilde{A}_2) = \frac{3}{8}$, e.t.c. Such that the resistances become $\mathscr{R}(\tilde{A}) = \frac{3}{8} \ln\left(1 + e^{\beta(1+\alpha)}\right)$, $\mathscr{R}(\tilde{B}) = \frac{1}{8} \ln\left(1 + e^{\beta}\right)$, $\mathscr{R}(\tilde{A}_1) = \frac{1}{8} \ln\left(1 + e^{\beta(1+\alpha)}\right)$, $\mathscr{R}(\tilde{A}_2) = \frac{1}{4} \ln\left(1 + e^{\beta(1+\alpha)}\right)$. The expected waiting time for each recurrent class can then be obtained by directly substituting for the respective resistance into the expression of Theorem 1.

If each of the eight agents were to interact globally, then the resistance of $\tilde{A}$ and $\tilde{B}$ would respectively be $\mathscr{R}(\tilde{A}) = \frac{1+\alpha}{2+\alpha} \ln\left(1 + e^{\beta(1+\alpha)}\right)$ and $\mathscr{R}(\tilde{B}) = \frac{1}{2+\alpha} \ln\left(1 + e^{\beta}\right)$. Which for $\alpha > 0$ are generally greater than those for the network topology of Figure 1. Similarly, when compared to the case in which the eight agents are arranged in a circle, in which case $\mathscr{R}(\tilde{A}) = \ln\left(1 + e^{\beta(1+\alpha)}\right)$ and $\mathscr{R}(\tilde{B}) = \frac{1}{n} \ln\left(1 + e^{\beta}\right)$, the resistance of the transition $\vec{A} \to \vec{B}$ is higher than for the case of

Figure 1. Hence the presence of such cohesive subgroups, though generally enhances the number of recurrent classes (some of which involve co-existence of choices), speeds up evolutions between them.

In general, it is possible to establish the most likely outcome (at least for binary coordination games) based on the the distribution of degrees in the population. Consider the binary coordination game in Table 1 and let

$$\eta_A = \frac{v(B,B) - v(A,B)}{v(B,B) - v(A,B) + v(A,A) - v(B,A)} \tag{12}$$

such that $\eta_B = 1 - \eta_A$. That is η_A is the fraction of neighbors that each agent requires to switch to A before doing so, and vise versa for η_B. Let $\lceil m \rceil$ be the smallest integer not less than m. It is easy to see that for any network G, each agent whose degree is such that $k_i = \lceil \eta_B k_i \rceil$ must be among those that form a threshold subset of agents required to switch to B for a global cascade to $\vec{B}$ to occur. The direct implication is that any arbitrary interaction structure for which the maximum degree $k_{max} = \lceil \eta_B k_{max} \rceil$, the state $\vec{A}$ is the most likely outcome in a long-run when the population size is sufficiently large. A typical example is the cyclic interaction above, whenever $\eta_B > \frac{1}{2}$.

4.3. Random networks

Most real world networks assume complex structures. Several authors have proposed various ways to capture the properties of such networks e.g. clustering coefficient and degree distribution. For our purpose, the degree distribution is the most suitable for capturing the effect of network topology on the diameter and average probabilities (hence resistance and expected waiting time) of inter-basin of attraction transitions. The degree distribution $P = \{p(k)\}_{k \geq 0}$ describes the number of agents in the population with a given degree. We denote by $\langle k \rangle$ for average degree. The following definitions related to the degree distribution are useful for the results that follow.

DEFINITION 2: *Let G and G' be two networks with respective degree distributions P and P'.*

(i) A degree distribution P is said to first order stochastically dominate (FOSD) P' if $\sum_{k=0}^{K} p(k) \leq \sum_{k=0}^{K} p'(k)$ for $1 \leq K \leq \infty$. Or for any non-decreasing function $f : \mathbb{R} \to \mathbb{R}$,

$$\sum_{k=0}^{\infty} f(k)p'(k) \leq \sum_{k=0}^{\infty} f(k)p(k).$$

(ii) A degree distribution P is said to second order stochastically dominate (SOSD) P' if for any non-decreasing concave function $f : \mathbb{R} \to \mathbb{R}$, $\sum_{k=0}^{\infty} f(k)p'(k) \leq \sum_{k=0}^{\infty} f(k)p(k)$.

(ii) A degree distribution P is said to be a mean preserving spread (MPS) of P' if P' SOSD P and that both have the same mean.

The following proposition provides the relationship between degree distributions and expected waiting time.

PROPOSITION 1: *Let G and G' be two networks with respective degree distributions P and P'. Let the underlying game be the binary coordination game of Table 1 in which A is the risk-dominant actions (i.e. $\eta_B > \frac{1}{2}$), and let $\mathscr{T}_G(\vec{B})$ denote the expected waiting time to exit $\tilde{B}$ given the network G. For a given value of η_B and revision rule,*

(i) if P FOSD (or SOSD) P', then $\mathscr{T}_G(\vec{B}) > \mathscr{T}_{G'}(\vec{B})$

(ii) if P MPS P' or P' FOSD P, then $\mathscr{T}_G(\vec{A}) > \mathscr{T}_{G'}(\vec{A})$

Proof. See Appendix A.2 □

The direct implication of Proposition 1 is that not only do sparsely connected networks (or presence of less connected agents) speed up evolution towards the state in which all agents play risk-dominant action, they also increase the expected waiting time to evolve away from it. Hence, for finite populations sparsely connected networks favor selection of risk-dominant equilibrium.

5. Convergence and diffusion rates

In this section, we provide convergence measures that can be employed to characterize other aspects of evolutionary game dynamics that are not directly captured by expected waiting time. More specifically, the convergence rate as a measure of the long-run behavior of the process and the diffusion rate as the measure of how fast an action, behavior or strategy diffuses across the population once its threshold value has been reached. The diffusion rate therefore captures the dynamics within a given basin of attraction once the process has entered its boundaries.

5.1. *Long-run behavior*

The definition for the convergence rate that we adopt is based on the notion that all finite state Markov chains attain heir stationary distribution at an exponential rate Seneta (1997). That is

$$\left|\left|P_\varepsilon^t(\mathbf{x}_{t_0},.) - \pi_\varepsilon\right|\right| \leq C r_\varepsilon^t \tag{13}$$

where C is some constant, $r < 1$, and $||.||$ is the total variation distance. The following proposition defines and provides a bound for the rate at which the process $(\mathbf{X}, P_\varepsilon)$ converges to its long-run stationary distribution.

Theorem 2: *Given any initial state $\mathbf{x}_{t_0}$ of the process $(\mathbf{X}, P_\varepsilon)$, the convergence rate $1 - r_\varepsilon$ of $(\mathbf{X}, P_\varepsilon)$ is given by*

$$1 - r_\varepsilon = \limsup_{t\to\infty} \left|\left|P_\varepsilon^t(\mathbf{x}_{t_0},.) - \pi_\varepsilon\right|\right|^{\frac{1}{t}} \geq \frac{1}{2}\left(\frac{1}{\mathscr{T}(\Omega^*)}\right)^2 \tag{14}$$

where $\Omega^* = \arg\min_{\Omega\in\mathbf{\Omega}} \mathscr{T}(\Omega)$

Proof. See Appendix A.3 □

Theorem 2 provide a lower bound for the convergence rate of an evolutionary process to stationarity in term of the maximum expected waiting time. Following from the results on the expected waiting time above, evolutionary processes on networks made up of cohesive subgroups generally attain their stationary distribution faster than global those under global interactions. In some cases in which the the network structure leads to almost global convergence to risk-dominant choice, such as the example of cyclic network structure given above, the convergence rate is faster global interactions that local interactions.

5.2. *Diffusion rate*

This section is devoted to deriving bounds for diffusion rate in terms of the parameters of the network topology and the underlying game. Given the normalized adjacency matrix $\mathscr{A}$ associated with the network of interactions, let $\rho(\mathscr{A}) = (\lambda_1, \cdots, \lambda_n)$ be its eigenvalue spectrum ordered in such away that $\lambda_1 = 1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Denote by Σ_ε for the an agent's individual transition matrix given an opponent's actions. That is let $\mathbb{P}(x_j|x_i)$ be the probability that a given player plays action $x_j \in X$ in the next period given that his opponent is playing $x_i \in X$ in the current period. Then Σ_ε is given by

$$\Sigma_\varepsilon = \begin{pmatrix} \mathbb{P}(x_1|x_1) & \mathbb{P}(x_2|x_1) & \cdots & \mathbb{P}(x_m|x_1) \\ \mathbb{P}(x_1|x_2) & \mathbb{P}(x_2|x_2) & \cdots & \mathbb{P}(x_m|x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{P}(x_1|x_m) & \mathbb{P}(x_2|x_m) & \cdots & \mathbb{P}(x_m|x_m) \end{pmatrix} \tag{15}$$

Let also $\rho(\Sigma_\varepsilon) = (\vartheta_1, \cdots, \vartheta_m)$ be the eigenvalue spectrum of Σ_ε. The following theorem provides bounds for diffusion rate in binary coordination games.

THEOREM 3: *Consider a set of n agents playing a binary coordination in Table 1 and let the evolutionary dynamics be that prescribed by (3). Then the diffusion rate $1 - r_D$ of an action, say A once its threshold has been reached is such that*

$$\lambda_2 \vartheta_2 \leq r_D \leq \lambda_2$$

Proof. See Appendix A.4 □

The second eigenvalue of Σ_ε, ϑ_2 is a function of the underlying payoff structure and the model of mistakes, which in the case of logit dynamics depends on β. A bound on λ_2 can be established through its relationship with the graph conductance $\phi(G)$. We give examples for specific graphs below.

EXAMPLE: The following relation between second eigenvalue of a network graph G, $\lambda_2(G)$ and $\phi(G)$ can be derived from Cheeger inequality: See Appendix A.5 for more detail concerning the relation plus derivations for the following examples.

$$\lambda_2(G) \leq 1 - \frac{\phi(G)^2}{2}. \tag{16}$$

(i) Complete network (G_{com}), a network structure in which every vertex is connected to every other vertex: $\lambda_2(G_{com}) \leq \frac{7}{8}$.

(ii) $1 - D$ cyclic network (G_{cyc}), a network in which vertices are arranged in a circle and every vertex is connected to two other neighboring vertices: $\lambda_2(G_{cyc}) \leq \frac{n^2-2}{n^2}$.

(iii) $2D$ $n \times n$ lattice network (G_{2D}), a lattice structure constructed with periodic boundary conditions such that each agent is connected to 4 neighbors: $\lambda_2(G_{2D}) \leq \frac{16n^2-1}{16n^2}$.

(iv) Random d-regular network (G_{d-r}), a network structure in which each of the n vertices is connected to d other vertices chosen at random: $\lambda_2(G_{d-r}) \leq \frac{7}{8}$.

(v) Newman's small world network (G_{nsw}), a network structure in which the mean shortest-path between nodes increases sufficiently slowly (logarithmically) as a function of the number of nodes in the network: $\lambda_2(G_{nsw}) = 1 - \mathcal{O}\left(\frac{c}{(\ln n)^2}\right)$, where c is a constant.

(vi) For a network with at least two cohesive subgroups, we have $\frac{7}{8} < \lambda_2(G) < 1$

Proof. See Appendix A.5 □

In general, densely connected and random networks have higher second largest eigenvalues compared to sparesely connected network. The direct implication of this result, together with the finding on expected waiting time, is that diffusion of strategic behavior or choices is faster in networks made up of cohesive subgroups. These cohesive subgroups should be small enough to favor shorter expected waiting time but large enough and sufficiently connected to favor higher diffusion rate within them.

6. Concluding remarks

This paper develops convergence measures for characterizing the short-run, medium-run and long-run behavior of stochastic evolutionary models. We provide tighter bounds for diffusion rate, expected waiting time and convergence rate as respective measures for the speed at which strategic behavior spreads across the population, the time for a transition between quasi-stable states and the the speed at which the process attains its invariant distribution. The bounds we derive do not rely on restrictive assumption concerning the model of mistakes, underlying game and interaction structure. Unlike the frameworks in the literature that impose assumptions on the model of mistakes, we find that the expected waiting time to exit the boundaries of a given basin of attraction depends both on its size and average probability of mistakes. The direct implication is that the quasi-stable states with the largest size of basin of attraction are not necessarily the most stable. This finding suggests the possibility of multiple long-run outcomes in models of stochastic evolution as opposed to the unique predictions always derived by employing stochastic stability as a solution concept. Our future work will focus on applying the results on expected waiting time to derive a general treatment of stochastic stability as an equilibrium selection method.

We have applied results from the three convergence measures to characterize strategic diffusion in networks. Several insights into the effect of the network topology on the diffusion of strategic behavior are established. For example we find that the expected waiting time under global interactions grows exponentially with population size while under local interactions it is independent of population size. But in addition to this well established result, we are able to demonstrate how the presence of cohesive subgroups in the network speeds up evolution between quasi-stable states. This effect is different from that of local interactions. We find that sparsely connected networks reinforce almost global stability and reduce the rate of inter-quasi-stable states transitions.

A. Appendix

A.1. Proof of Theorem 1

From the definition of expected waiting time, we have that

$$\mathbb{E}\left[\min\left\{t|\ n(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)\geq d(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)n\right\}\right]=\frac{1}{\mathbb{P}(n(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)\geq d(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)n)} \tag{A.1}$$

The objective is therefore to place a bound on $\mathbb{P}(n(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)\geq d(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)n$. The derivation follows a combinatorial argument, where we first transform individual transition probabilities into Boolean random variables. That is, let $\mathbb{P}_i(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)=\mathbb{P}_i(x_{t+1}=y|\mathbf{x}_t=\mathbf{x})$ for $y\in\mathbf{y}\in\Omega_j\in\tilde{\mathbf{x}}_j$ and $y\notin\mathbf{x}\in\Omega_i\in\tilde{\mathbf{x}}_i$. Define a parameter $p\in[0,1]$ such that if $\mathbb{P}_i(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)\geq p$ agent i chooses $y\in\mathbf{y}$ or else he does not. This leads to a random variable denoted by I_i, which is equal to one if i chooses y and zero otherwise. Let $I=(I_1,\cdots,I_n)$ be the realization of I_i for all $i\in N$. From the definition of $n(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)$, we then rephrase our problem as the case of bounding $\mathbb{P}(\sum_{i=1}^n I_i\geq d(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)n)$.

Now, consider the problem of a binomial independent sampling over the vector I, $Bin(n,\sigma)$, such that with probability σ, I_i is picked and with $1-\sigma$ it is not. Denote the n-dimensional vector generated by $Bin(n,\sigma)$ by $\mathbf{u}=(u_1,\cdots,u_n)$, where $\mathbb{P}(u_i=1)=\sigma$ and $\mathbb{P}(u_i=0)=1-\sigma$. We can then regard the problem of bounding $\mathbb{P}(\sum_{i=1}^n I_i\geq d(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)n)$ as determining the probability of "efficiently" finding a subset $S\subseteq N$ of at least $d(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)n$ players all of whom simultaneously switch to play y. Define an event $\forall_{i\in S}I_i=1$; that is all members of S choose y, and consequently $\mathbb{P}(\forall_{i\in S}I_i=1)$ is the probability that all $i\in S$ choose y. We can then define the following conditional relation,

$$\mathbb{P}\left(\sum_{i=1}^n I_i\geq d(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)n\right)\leq\frac{\mathbb{E}[\forall_{i\in S}I_i=1]}{\mathbb{E}[\forall_{i\in S}I_i=1|\sum_{i=1}^n I_i\geq d(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)n]}, \tag{A.2}$$

where the expectations are taken over the vector $\mathbf{u}$. Since the elements of $\mathbf{u}$ are a result of independent sampling, we have

$$\mathbb{E}[\forall_{i\in S}I_i=1]\leq\sum_{S\subseteq N}\left(\sigma^{\#S}(1-\sigma)^{n-\#S}\prod_{i\in S}\mathbb{P}_i(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)\right)=\mathbb{E}_{\forall u_i\in\mathbf{u}}\left[\prod_{i=1}^n\left(\mathbb{P}_i(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)|u_i\right)\right],$$

where the first inequality follows from the fact that $\mathbb{P}(\forall_{i\in S}I_i=1)\leq\prod_{i\in S}\mathbb{P}_i(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)$. It follows that,

$$\mathbb{E}[\forall_{i\in S}I_i=1]\leq\prod_{i=1}^n\mathbb{E}_{u_i}\left[\mathbb{P}_i(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)|u_i\right]=\prod_{i=1}^n\left(\sigma\mathbb{P}_i(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)+1-\sigma\right) \tag{A.3}$$

If we define $\mathbb{P}_A(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)=\frac{1}{n}\sum_{i=1}^n\mathbb{P}_i(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)$ as the arithmetic average of all $\mathbb{P}_i(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)$, then from the convex relation between the logarithms of the arithmetic and geometric means,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n\ln\left(\sigma\mathbb{P}_i(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)+1-\sigma\right)\leq\ln\left(\sigma\mathbb{P}_A(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)+1-\sigma\right)\\ &\mathbb{E}[\forall_{i\in S}I_i=1]\leq\prod_{i=1}^n\left(\sigma\mathbb{P}_i(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)+1-\sigma\right)\leq\left(\sigma\mathbb{P}_A(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)+1-\sigma\right)^n\end{aligned} \tag{A.4}$$

To obtain the bound for $\mathbb{E}[\forall_{i\in S}I_i=1|\sum_{i=1}^n I_i\geq d(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)n]$, recall that $1-\sigma$ is the probability that $u_i=0$. We also note that if at least $d(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)n$ of the elements of $\mathbf{u}$ are ones, then there are at

most $n - d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)n$ zeros, that is at most $n - d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)n$ agents are not in set S. It follows that

$$\mathbb{E}\left[\forall_{i\in S} I_i = 1 | \sum_{i=1}^{n} I_i \geq d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)n\right] \geq (1-\sigma)^{(1-d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j))n} \tag{A.5}$$

Equations (A.4) together with (A.5) when substituted into (A.2) yield,

$$\mathbb{P}\left(\sum_{i=1}^{n} I_i \geq d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)n\right) \leq \left(\frac{\left(\sigma \mathbb{P}_A(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) + 1 - \sigma\right)}{(1-\sigma)^{(1-d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j))}}\right)^n, \tag{A.6}$$

If we choose σ that optimizes the quantity $g = \frac{\left(\sigma \mathbb{P}_A(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) + 1 - \sigma\right)}{(1-\sigma)^{(1-d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j))}}$ (by equating the derivative to zero and solving for σ) and substituting back gives

$$\mathbb{P}\left(\sum_{i=1}^{n} I_i \geq d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)n\right) \leq \left(\left(\frac{\mathbb{P}_A(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)}{d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)}\right)^{d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)} \left(\frac{1 - \mathbb{P}_A(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)}{1 - d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)}\right)^{1-d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)}\right)^n \tag{A.7}$$

Re-expressing (A.7) in exponential form results to

$$\mathbb{P}(n(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) \geq d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)n) \leq e^{n\left[d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)\ln\left(\frac{\mathbb{P}_A(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)}{d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)}\right) + (1-d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j))\ln\left(\frac{1-\mathbb{P}_A(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)}{1-d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)}\right)\right]} \tag{A.8}$$

Note that each $\mathbb{P}_i(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) = \mathscr{P}_i(x_{t+1} = y | \mathbf{x}_t = \mathbf{x} \in \Omega_i) \ll 1$, hence $\ln(1 - \mathbb{P}_A(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)) \ll \ln(1 - d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j))$ such that

$$\mathbb{P}(n(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) \geq d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)n) \leq e^{-n[f(d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)) - d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)\ln(\mathbb{P}_A(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j))]} \tag{A.9}$$

where $f(d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)) = d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)\ln\left(\frac{d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)}{1-d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)}\right) + \ln\left(d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)\right)$. Recall that the cost of the transition $\tilde{\mathbf{x}}_i \to \tilde{\mathbf{x}}_j$ is

$$c(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) = -d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)\ln\left(\mathbb{P}_A(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)\right).$$

It then follows from the definition of the resistance $\mathscr{R}(\tilde{\mathbf{x}}_i)$ of $\tilde{\mathbf{x}}_i$ that

$$\begin{aligned}
\mathscr{T}(\Omega_i) &= \min_{\tilde{\mathbf{x}}_j \neq \tilde{\mathbf{x}}_i} \mathbb{E}\left[\min\left\{t |\ n(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) \geq d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)n\right\}\right] \\
&\leq \min_{\tilde{\mathbf{x}}_j \neq \tilde{\mathbf{x}}_i} e^{n[f(d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)) + c(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)]} \\
&= e^{-n[f(d(\tilde{\mathbf{x}}_i)) + \mathscr{R}(\tilde{\mathbf{x}}_i)]}
\end{aligned} \tag{A.10}$$

where $d(\tilde{\mathbf{x}}_i) = \arg\min_{d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)} c(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)$.

A.2. Proof of Proposition 1

(i) For a given value of $\eta_A < \frac{1}{2}$ the quantity $d(\tilde{B}, \tilde{A})$ is an increasing function of network connectivity. For example when the network is complete network (global interactions) $d(\tilde{B}, \tilde{A}) = \eta_B$ and for cyclic interactions $d(\tilde{B}, \tilde{A}) = \frac{1}{n}$. We thus have that $d(\tilde{B}, \tilde{A}) := f(\langle k \rangle)$, where f is an increasing function. Since P FOSD (or SOSD) P' implies that $\langle k \rangle > \langle k' \rangle$, it follows that $d(\tilde{B}, \tilde{A})$ under G is greater than $d'(\tilde{B}, \tilde{A})$ under G'. When agents are matched randomly and uniformly the average probabilities $\mathbb{P}_A(\tilde{B}, \tilde{A})$ are identical for both networks. Implying that $\mathscr{T}_G(\vec{B}) > \mathscr{T}_{G'}(\vec{B})$.

(ii) Given $\frac{1}{2} < \eta_B < 1$, all agents for whom $k_i = \lceil \eta_B k_i \rceil$ require all their neighbors to switch to B before doing so. Denote by $N_B(G)$ for the subset of agents for whom $k_i = \lceil \eta_B k_i \rceil$ given the network G. That is

$$N_B(G) = \{i \in N :\ k_i = \lceil \eta_B k_i \rceil\}.$$

Denote k_{max} for the maximum k for which $k = \lceil \eta_B k \rceil$. We then have that

$$n_A(G) = \#N_A(G) = \sum_{k=0}^{k_{max}} p(k)$$

such that

$$d_G(\tilde{A}, \tilde{B}) \geq \frac{n_A(G)}{n} = \frac{\sum_{k=0}^{k_{max}} p(k)}{n}$$

It then implies that for P' FOSD P or P MPS P', there exists a value of k_{max} sufficiently large such that

$$d_G(\tilde{A}, \tilde{B}) \geq \frac{\sum_{k=0}^{k_{max}} p(k)}{n} > \frac{\sum_{k=0}^{k_{max}} p'(k)}{n} \leq d_{G'}(\tilde{A}, \tilde{B}).$$

Implying that $\mathscr{T}_G(\vec{A}) > \mathscr{T}_{G'}(\vec{A})$.

A.3. Proof of Theorem 2

The proof follows from linear algebra (see for example Jerrum and Sinclair (1989) for general Markov chains). We drop the subscript on P_ε for notational convenience. Recall that P_ε is reversible.

$$\pi_\varepsilon(\mathbf{x})P_\varepsilon(\mathbf{x}, \mathbf{y}) = \pi_\varepsilon(\mathbf{y})P_\varepsilon(\mathbf{y}, \mathbf{x}) \qquad \forall \mathbf{x}, \mathbf{y} \in \mathbf{X} \tag{A.11}$$

We can define an equivalent symmetric matrix M such that

$$M(\mathbf{x}, \mathbf{y}) = \sqrt{\frac{\pi_\varepsilon(\mathbf{x})}{\pi_\varepsilon(\mathbf{y})}} P_\varepsilon(\mathbf{x}, \mathbf{y}) \tag{A.12}$$

Denote the diagonal matrix with elements $\pi_\varepsilon(\mathbf{x})$ by D such that

$$M = D^{\frac{1}{2}} P_\varepsilon D^{-\frac{1}{2}} \tag{A.13}$$

Let $h_i = D^{-\frac{1}{2}}\mathbf{u}_i$, (where $\mathbf{u}_i$'s are eigenfunctions of M) it follows that h_i is an eigenfunction of P_ε corresponding to eigenvalue μ_i, and that h_i's are orthonormal with respect to π_ε. That is

$$P_\varepsilon h_i = P_\varepsilon D^{-\frac{1}{2}}\mathbf{u}_i = D^{-\frac{1}{2}}\left(D^{\frac{1}{2}} P_\varepsilon D^{-\frac{1}{2}}\right)\mathbf{u}_i = D^{-\frac{1}{2}} M \mathbf{u}_i = D^{-\frac{1}{2}}\mu_i \mathbf{u}_i = \mu_i h_i. \tag{A.14}$$

To prove orthonormality of the functions h_i's with respect to π_ε, its useful to take note of the following definition of the *inner product*. Let $\langle ., . \rangle$ denote the inner product on $\mathbb{R}^{\mathbf{X}}$, that is

$$\langle h, g \rangle = \sum_{\mathbf{x} \in \mathbf{X}} h(\mathbf{x}) g(\mathbf{x}), \tag{A.15}$$

then we can define the inner product with respect to the distribution π_ε as

$$\langle h, g \rangle_{\pi_\varepsilon} = \sum_{\mathbf{x} \in \mathbf{X}} h(\mathbf{x}) g(\mathbf{x}) \pi_\varepsilon(\mathbf{x}). \tag{A.16}$$

Let $\delta_{i,j}$ denote the Dirac delta function (that is $\delta_{i,j} = 1$ if and only if $i = j$), then

$$\delta_{i,j} = \langle \mathbf{u}_i, \mathbf{u}_j \rangle = \left\langle D^{\frac{1}{2}} h_i, D^{\frac{1}{2}} h_j \right\rangle = \sum_{\mathbf{x} \in \mathbf{X}} \pi_\varepsilon(\mathbf{x})^{\frac{1}{2}} h(\mathbf{x}) \pi_\varepsilon(\mathbf{x})^{\frac{1}{2}} g(\mathbf{x}) = \sum_{\mathbf{x} \in \mathbf{X}} h(\mathbf{x}) \pi_\varepsilon(\mathbf{x}) g(\mathbf{x}) = \langle h, g \rangle_{\pi_\varepsilon}.$$

Note that $P_\varepsilon^t(\mathbf{x},\mathbf{y})$ is the $\mathbf{x},\mathbf{y}$ element of P_ε^t, implying that $P_\varepsilon^t(\mathbf{x},\mathbf{y}) = (P_\varepsilon^t\delta_\mathbf{y})(\mathbf{x})$; where $\delta_\mathbf{y}(\mathbf{x})$ is a Dirac function assuming the value of unity for $x = y$ and zero otherwise. Notice also that $\delta_\mathbf{y}$ belongs to the inner product space $\mathbb{V} = (\mathbb{R}^\mathbf{X}, \langle .,. \rangle_{\pi_\varepsilon})$, and since the set $\{h_1, \cdots, h_{|\mathbf{X}|}\}$ is an orthonormal basis of $\mathbb{V}$, then $\delta_\mathbf{y}$ can be written via basis decomposition as

$$\delta_\mathbf{y} = \sum_{i=1}^{\eta} \langle \delta_\mathbf{y}, h_i \rangle_{\pi_\varepsilon} h_i = \sum_{i=1}^{\eta} h_i(\mathbf{y})\pi_\varepsilon(\mathbf{y})h_i \tag{A.17}$$

where η is the cardinality of $\mathbf{X}$. Substituting (A.17) and $P_\varepsilon^t h_i = \mu_i^t h_i$ gives $P_\varepsilon^t(\mathbf{x},\mathbf{y})$ as

$$P_\varepsilon^t(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{\eta} h_i(\mathbf{y})\pi_\varepsilon(\mathbf{y})\mu_i^t h_i(\mathbf{x}) = \pi_\varepsilon(\mathbf{y}) + \sum_{i=2}^{\eta} h_i(\mathbf{y})\pi_\varepsilon(\mathbf{y})\mu_i^t h_i(\mathbf{x}) \tag{A.18}$$

It then follows that,

$$\begin{aligned}
\limsup_{t\to\infty} \left|\left| P_\varepsilon^t(\mathbf{x}_0,.) - \pi_\varepsilon(.) \right|\right|^{1/t} &= \limsup_{t\to\infty} \left[\frac{1}{2} \sum_{\mathbf{y}\in\mathbf{X}} \sum_{i=2}^{\eta} |h_i(\mathbf{y})\pi_\varepsilon(\mathbf{y})h_i(\mathbf{x}_0)\mu_i^t| \right]^{1/t} \\
&= |\mu_2| \limsup_{t\to\infty} \left[\frac{1}{2} \sum_{\mathbf{y}\in\mathbf{X}} \left(|h_2(\mathbf{y})\pi_\varepsilon(\mathbf{y})h_2(\mathbf{x}_0)| + \frac{1}{|\mu_2|^t} \sum_{i=3}^{\eta} |h_i(\mathbf{y})\pi_\varepsilon(\mathbf{y})h_i(\mathbf{x}_0)\mu_i^t| \right) \right]^{1/t} \\
&= |\mu_2| \qquad (A.19)
\end{aligned}$$

Recall that each basin of attraction $\tilde{\mathbf{x}} \subset \mathbf{X}$ and sets of basins of attraction $W \subset \tilde{\mathbf{X}}$ induce a partition on P_ε. Denote by $\phi(P_\varepsilon, W)$ for the conductance of the partition on P_ε induced by W. That is

$$\phi(P_\varepsilon, W) = \frac{1}{\sum_{\mathbf{x}\in W} \pi(\mathbf{x})} \sum_{\mathbf{x}\in W} \sum_{\mathbf{y}\in W^c} \pi(\mathbf{x})P_\varepsilon(\mathbf{x},\mathbf{y}). \tag{A.20}$$

where W^c is the complement of W. The conductance of P_ε is then defined by

$$\phi(P_\varepsilon) = \min_{\substack{W\subset\mathbf{X} \\ \sum_{\mathbf{x}\in W}\pi(\mathbf{x})\le\frac{1}{2}}} \phi(P_\varepsilon, W) \tag{A.21}$$

If the basins of attraction are of sufficiently large sizes, then

$$\phi(P_\varepsilon) = \min_{\substack{\tilde{\mathbf{x}}_i\subset\tilde{\mathbf{X}} \\ \sum_{\mathbf{x}\in\tilde{\mathbf{x}}_i}\pi(\mathbf{x})\le\frac{1}{2}}} \phi(P_\varepsilon, \tilde{\mathbf{x}}_i) \tag{A.22}$$

Given the definition of conductance of P_ε, the following relationship holds (Jerrum and Sinclair, 1989)

$$\mu_2 \le 1 - \frac{\phi(P_\varepsilon)^2}{2} \tag{A.23}$$

From (A.20) and letting $\pi_\varepsilon(\tilde{\mathbf{x}}_i) = \sum_{\mathbf{x}\in\tilde{\mathbf{x}}_i} \pi(\mathbf{x})$, we then have that

$$\begin{aligned}
\phi(P_\varepsilon, \tilde{\mathbf{x}}_i)) &= \frac{1}{\pi_\varepsilon(\tilde{\mathbf{x}}_i)} \sum_{\mathbf{x}\in\tilde{\mathbf{x}}_i} \pi(\mathbf{x}) \sum_{\mathbf{y}\in\tilde{\mathbf{x}}_i^c} P_\varepsilon(\mathbf{x},\mathbf{y}) \\
&\ge \min_{\mathbf{x}\in\tilde{\mathbf{x}}_i} \sum_{\tilde{\mathbf{x}}_j\neq\tilde{\mathbf{x}}_i} \sum_{\mathbf{y}\in\tilde{\mathbf{x}}_i^c} P_\varepsilon(\mathbf{x},\mathbf{y}) \\
&\ge \sum_{\tilde{\mathbf{x}}_j\neq\tilde{\mathbf{x}}_i} \mathbb{P}(n(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j) \ge d(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)n) \\
&\ge \max_{\tilde{\mathbf{x}}_i\neq\tilde{\mathbf{x}}_i} \mathbb{P}(n(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j) \ge d(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)n) = \frac{1}{\mathscr{T}(\Omega_i)}
\end{aligned}$$

Implying that

$$\phi(P_\varepsilon) = \min_{\tilde{\mathbf{x}}_i \in \tilde{\mathbf{X}}} \phi(P_\varepsilon, \tilde{\mathbf{x}}_i)) = \max_{\Omega_i \in \mathbf{\Omega}} \frac{1}{\mathscr{T}(\Omega_i)} \tag{A.24}$$

A.4. Proof of Theorem 3

We begin by proving the following lemma.

LEMMA 1: *Let μ_3 and μ_4 be the third and fourth largest eigenvalues of P_ε. Then the diffusion rate r_D is such that $\mu_4 \leq r_D \leq \mu_3$.*

Proof. Recall that the partition of the state space $\mathbf{X}$ into basins of attractions induced by the regular perturbation on P, makes P_ε nearly decomposable. In the case of binary action set say $X = \{A, B\}$, P_ε is nearly decomposable into diagonal block matrices $l = \{1, 2\}$ describing transitions within elements of the basins of attractions $\tilde{A}$ and $\tilde{B}$. We can write eigenvalue spectrum of P_ε as $\rho(P_\varepsilon) = (\mu_{1_1}, \mu_{2_1}, \cdots, \mu_{n_1}, \mu_{1_2}, \cdots, \mu_{n_2})$. Where n_l is the size of block matrix l. μ_{1_1} is the largest eigenvalue of the first diagonal block matrix also corresponding to the leading eigenvalue of the the entire matrix. μ_{n_1} is the least eigenvalue of the first block matrix. μ_{1_2} and μ_{n_2} are the leading and least eigenvalues of the second block matrix.

Let $\mathbf{r}_{j_l}$ and $\mathbf{w}_{\mathbf{j_l}}$ be the right and left eigenvectors corresponding to the eigenvalue μ_{j_l}. By expressing P_ε in its eigendecomposition form, the dynamic process can be expressed as

$$\mathbf{q}_t = \mathbf{q}_0 \mathbf{r}_{1_1} \mathbf{w}_{1_1}^T + \sum_{j=2}^{n_1} \mu_{j_1}^t \mathbf{q}_0 \mathbf{r}_{j_1} \mathbf{w}_{j_1}^T + \mu_{1_2}^t \mathbf{q}_0 \mathbf{r}_{1_2} \mathbf{w}_{1_2}^T + \sum_{j=2}^{n_2} \mu_{j_2}^t \mathbf{q}_0 \mathbf{r}_{j_2} \mathbf{w}_{j_2}^T \tag{A.25}$$

First note that $\mu_{1_2} > \mu_{j_l}$ for all $j \geq 2$. Implying that the convergence rate of the entire adoptive process in a long-run according to Proposition 2, is given by

$$r = \limsup_{t \to \infty} \left|\left| P_\varepsilon^t(\mathbf{x}_{t_0}, .) - \pi \right|\right|^{\frac{1}{t}} = \mu_{1_2}.$$

Since the leading eigenvalues of each block matrix are such that $\mu_{1_1} = 1$ and μ_{1_2} is very close to unity, in a medium-run the varying terms in (A.25) are the second and fourth, and the first and third terms stay fairly constant. Once the process enters a given diagonal block matrix (boundaries of a given basin of attraction) it attains a quasi-stationary distribution. The quasi-stationary distribution places most weight on the most stable outcome (quasi-stable state) within the basin of attraction, which in the case of binary coordination games would $\vec{A}$ and $\vec{B}$. The rate at which $(\mathbf{X}, P_\varepsilon)$ converges to its quasi-stationary distribution within a given basin of attraction is thus approximately equal to the rate at which a given action (say A) gets globally adopted through best response.

Denote by π_A and π_B for the quasi-stationary distribution within the basin of attraction of $\vec{A}$ and $\vec{B}$ respectively. The convergence rates for each block matrix are then

$$r = \limsup_{t \to \tau} \left|\left| P_\varepsilon^t(\mathbf{x}_{t_0}, .) - \pi_1 \right|\right|^{\frac{1}{t}} = \mu_{2_1}.$$

$$r = \limsup_{t \to \tau} \left|\left| P_\varepsilon^t(\mathbf{x}_{t_0}, .) - \pi_2 \right|\right|^{\frac{1}{t}} = \mu_{2_2}$$

where τ is the period at which π_A and π_B are attained. It then follows that

$$\mu_4 = \mu_{2_2} \leq r_D \leq \mu_{2_1} = \mu_3.$$

□

We can derive an approximation for the relationship between μ_3 and μ_4 with the parameters of the underlying game and interaction structure by considering the following linearized evolutionary process. The transition matrix P_ε can be linearized by acting on it an operator called the *event matrix* (Asavathiratham, 2001). Denote the event matrix and its Moore-Penrose pseudo-inverse by Ψ and Ψ^{-1} respectively. Ψ is derived by stacking into rows all possible realizations or state vectors written in *basis vector* form. For example in the case of two players and binary action set in which there are 4 possible realizations, the event matrix becomes

$$\Psi = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \tag{A.26}$$

where the first row corresponds to the state in which both players play action A, and so forth. For any finite number of players and action set, the following equivalence holds.

$$P_\varepsilon \Psi = \Psi \Pi_\varepsilon \tag{A.27}$$

where Π_ε is the linearized form of P_ε, a transition matrix defined as

$$\Pi_\varepsilon = \mathscr{A}^T \otimes \Sigma_\varepsilon \tag{A.28}$$

Given Π_ε and P_ε, the following lemma holds.

LEMMA 2: *Let* $\rho(\Pi_\varepsilon) = \theta_1, \cdots, \theta_{2n}$, $\rho(\mathscr{A}) = (\lambda_1, \cdots, \lambda_n)$ *and* $\rho(\Sigma_\varepsilon) = (\vartheta_1, \vartheta_2)$ *denote the eigenvalues of* Π_ε, $\rho(\mathscr{A})$ *and* $\rho(\Sigma_\varepsilon)$ *respectively. Then*

1. *If* μ_1 *and* θ_1 *are the unique largest eigenvalues of* P_ε *and* Π_ε *respectively, then* $\mu_1 = \theta_1 = 1$.
2. $\rho(\Pi_\varepsilon) = \lambda_i \vartheta_j \ \forall \lambda_i \in \rho(\mathscr{A}),\ \vartheta_j \in \rho(\Sigma_\varepsilon)$

Proof. Let $\mathbf{f}_i$ be the right eigenvector of P_ε corresponding to eigenvalue μ_i, and recall that the relation $P_\varepsilon \Psi = \Psi \Pi_\varepsilon$. It follows that

$$P_\varepsilon \Psi \mathbf{f}_1 = \Psi \Pi_\varepsilon \mathbf{f}_1$$

where $\mathbf{f}_1$ is the eigenvector corresponding to eigenvalue μ_1. Since $\mu_1 = 1$, we thus have $P_\varepsilon \Psi \mathbf{f}_1 = \Psi \mathbf{f}_1$, which is true if and only if $\Pi_\varepsilon \mathbf{f}_1 = \mathbf{f}_1$. That is $\theta_1 = \mu_1 = 1$.

For part 2 of the lemma see Horn and Johnson (1990, page 245, Theorem 4.2.12) □

Since

$$q_t \Psi = q_0 P_\varepsilon^t \Psi = q_0 \Psi \Pi_\varepsilon^t \tag{A.29}$$

It follows that $\mu_4 \approx \lambda_2 \vartheta_2$ and $\mu_3 \approx \lambda_2$.

A.5. Proofs for Example

Let $G = (n, E)$ be a graph or network of n vertices. Denote by S a subset of n and by $e(S, n-S)$ as the number of interactions (for a weighted network graph its is the sum of weighted interactions)

between S and its complement $n-S$. Also let $d(S)$ denote the total degree of subset S. Then the conductance of G

$$\phi(G) = \min_{S,\#S\leq\frac{n}{2}} \frac{e(S,n-S)}{d(S)} \tag{A.30}$$

where $\#S$ denotes the size of the set. For regular network graphs (in which all vertices have the same degree), it is shown by (Alon and Milman, 1985) that

$$\lambda_2(G) \leq 1 - \frac{\phi(G)^2}{2}. \tag{A.31}$$

For a complete graph, since every vertex is connected to every other vertex, we have that every vertex in S is connected to all other vertices in $n-S$. This implies that $e(S,n-S) = \#S\times\#(n-S) = \#S\times(n-\#S)$, and $d(S) = n\times\#S$ such that

$$\phi(G_{com}) = \min_{S,\#S\leq\frac{n}{2}} \frac{\#S\times(n-\#S)}{n\times\#S} \geq \frac{1}{2}, \tag{A.32}$$

where the last inequality follows from the fact that $\#S \leq \frac{n}{2}$. We thus have that

$$\lambda_2(G_{com}) \leq \frac{7}{8} \tag{A.33}$$

In the case of a $1-D$ cyclic network, $e(S,n-S) = 2$, and $d(S) = 2\times\#S$ such that

$$\phi(G_{cyc}) = \min_{S,\#S\leq\frac{n}{2}} \frac{2}{2\times\#S} \geq \frac{2}{n}. \tag{A.34}$$

Hence $\lambda_2(G_{cyc}) \leq \frac{n^2-2}{n^2}$.

$2D$ network: Let the composition of S be chosen in such a way that the peripheral vertices (vertices at the perimeter or boundary of S) contain approximately one edge each connecting it to the set $n-S$. Since it is a 2-dimensional structure there should be approximately $\sqrt{\#S}$ vertices forming such a boundary. This implies that $e(S,n-S) \approx \sqrt{\#S}$, and $d(S) = 4\times\#S$ such that

$$\phi(G_{2D}) = \min_{S,\#S\leq\frac{n^2}{2}} \frac{\sqrt{\#S}}{4\times\#S} \geq \frac{\sqrt{2}}{4n}. \tag{A.35}$$

where the last inequality follows from the fact that $\sqrt{\#S} \leq \sqrt{\frac{n^2}{2}}$. It follows that

$$\lambda_2(G_{2D}) \leq \frac{16n^2-1}{16n^2}.$$

Random d-regular network: Since for each vertex the vertices to which it is connected are chosen at random, and that the maximum size of S is $\frac{n}{2}$, then a typical vertex in S is connected to approximately $\frac{d\times\#(n-S)}{n}$ other vertices in $n-S$ such that $e(S,n-S) \approx \frac{d\times\#S\#(n-S)}{n}$. We thus have

$$\phi(G_{d-r}) = \min_{S,\#S\leq\frac{n}{2}} \frac{\frac{d\times\#S\#(n-S)}{n}}{d\times\#S} \geq 1. \tag{A.36}$$

$$\lambda_2(G_{d-r}) \leq \frac{7}{8} \tag{A.37}$$

For Newman's small world networks see Durrett (2006).

Recall the definition of a cohesive subgroups.

$$\frac{d(l, n-l)}{d(l)} \leq 1-\delta, \qquad \frac{1}{2} < \delta < 1 \tag{A.38}$$

Equation (A.38) together with (A.30) and (A.31) imply

$$\lambda_2 = 1 - \frac{1}{2}(1-\delta)^2, \qquad \frac{1}{2} < \delta < 1 \tag{A.39}$$

such that for $\delta = \frac{1}{2}$, $\lambda_2 = \frac{7}{8}$, and $\delta = 1$, $\lambda_2 = 1$.

Acknowledgments

This study has benefited from comments and suggestions of Julio Davila, Arkadi Predtetchinsk, Antoine Mandel, François Lafond, Hippolyte d'Albis, Giorgio Triulzi and Robin Cowan. We also gratefully acknowledge the fruitful discussions with seminar participants at The Centre d'Economie de La Sorbonne and Paris School of Economics, University of Cergy Pontoise THEMA, Maastricht Lecture Series in Economics and Cournot seminars at Bureau d'Economie Théorique et Appliquée (BETA), University of Strasbourg. This research was supported by UNU-MERIT. The usual disclaimer applies.

References

Akerlof, G. A. (1997). Social distance and social decisions. *Econometrica 65*(5), 1005–1028.

Alon, N. and V. D. Milman (1985). λ_1, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B 38*(1), 73 – 88.

Alós-Ferrer, C. and S. Weidenholzer (2008). Contagion and efficiency. *Journal of Economic Theory 143*(1), 251–274.

Asavathiratham, C. (2001). *The influence model: a tractable representation for the dynamics of networked Markov chains*. Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, Ph.D Thesis.

Bergin, J. and B. L. Lipman (1996). Evolution with state-dependent mutations. *Econometrica 64*(4), 943–56.

Binmore, K. and L. Samuelson (1997). Muddling through: Noisy equilibrium selection. *Journal of Economic Theory 74*(2), 235–265.

Blume, L. E. (1995). The statistical mechanics of best-response strategy revision. *Games and Economic Behavior 11*(2), 111–145.

Brock, W. A. and S. N. Durlauf (2001). Discrete choice with social interactions. *Review of Economic Studies 68*(2), 235–60.

Durrett, R. (2006). *Random Graph Dynamics*. Cambridge Series on Statistical and Probabilistic Mathematics. Cambridge University Press.

Ellison, G. (1993). Learning, local interaction, and coordination. *Econometrica 61*(5), 1047–1071.

Ellison, G. (2000). Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution. *Review of Economic Studies 67*(1), 17–45.

Foster, D. and P. Young (1990). Stochastic evolutionary game dynamics. *Theoretical population biology 38*(2), 219–232.

Glaeser, E. L. and J. A. Scheinkman (2001). Non-market interactions. Technical report.

Horn, R. A. and C. R. Johnson (1990). *Matrix Analysis*. Cambridge University Press.

Huck, S., D. Kübler, and J. Weibull (2012). Social norms and economic incentives in firms. *Journal of Economic Behavior & Organization 83*(2), 173 – 185.

Jerrum, M. and A. Sinclair (1989). Approximating the permanent. *SIAM J. Comput. 18*, 1149–1178.

Kandori, M., G. J. Mailath, and R. Rob (1993). Learning, mutation, and long run equilibria in games. *Econometrica 61*(1), 29–56.

Kandori, M. and R. Rob (1998). Bandwagon effects and long run technology choice. *Games and Economic Behavior 22*(1), 30–60.

Lee, I. H., A. Szeidl, and A. Valentinyi (2003). Contagion and State Dependent Mutations. *The B.E. Journal of Theoretical Economics 3*(1), 1–29.

Lelarge, M. (2012). Diffusion and cascading behavior in random networks. *Games and Economic Behavior 75*(2), 752–775.

Montanari, A. and A. Saberi (2010). The spread of innovations in social networks. *Proceedings of the National Academy of Sciences 107*(47), 20196–20201.

Morris, S. (2000). Contagion. *Review of Economic Studies 67*(1), 57–78.

Nöldeke, G. and L. Samuelson (1997). A dynamic model of equilibrium selection in signaling markets. *Journal of Economic Theory 73*(1), 118–156.

Sandholm, W. (2010). *Population Games and Evolutionary Dynamics*. Economic Learning and Social Evolution. MIT Press.

Sandholm, W. H. (2001). Almost global convergence to p-dominant equilibrium. *International Journal of Game Theory*, 107–116.

Seneta, E. (1997). *Non-negative Matrices and Markov Chains*. Springer Series in Statistics. Springer.

Staudigl, M. (2012). Stochastic stability in asymmetric binary choice coordination games. *Games and Economic Behavior 75*(1), 372 – 401.

Vega-Redondo, F. (1997). The evolution of walrasian behavior. *Econometrica 65*(2), 375–384.

Young, H. P. (1993). The evolution of conventions. *Econometrica 61*(1), 57–84.

Young, P. H. (2011). The dynamics of social innovation. *Proceedings of the National Academy of Sciences 108*(Supplement 4), 21285–21291.

The UNU-MERIT WORKING Paper Series

2014-01 *The medium-term effect of R&D on firm growth* by Marco Capasso, Tania Treibich and Bart Verspagen
2014-02 *Diverse and uneven pathways towards transition to low carbon development: The case of diffusion of solar PV technology in China*Michiko Iizuka
2014-03 *User innovators and their influence on innovation activities of firms in Finland* by JariKuusisto, MerviNiemi and Fred Gault
2014-04 *Migration, remittances and household welfare in Ethiopia* by Lisa Andersson
2014-05 *Path-breaking directions of nanotechnology-based chemotherapy and molecular cancer therapy* by Mario Coccia and Lili Wang
2014-06 *Poor trends - The pace of poverty reduction after the Millennium Development Agenda*Richard Bluhm, Denis de Crombrugghe, Adam Szirmai
2014-07 *Firms' adoption of international standards: Evidence from the Ethiopian floriculture sector* by MuluGebreeyesu
2014-08 *School choice, segregation, and forced school closure*by Cheng Boon Ong andKristof De Witte
2014-09 *Gender difference in support for democracy in Sub-Saharan Africa: Do social institutions matter?*by MatyKonte
2014-10 *Why are women less democratic than men? Evidence from Sub-Saharan African countries* by Cecilia García-Peñalosa and MatyKonte
2014-11 *Tipping points? Ethnic composition change in Dutch big city neighbourhoods* by Cheng Boon Ong
2014-12 *Technology life cycle and specialization patterns of latecomer countries. The case of the semiconductor industry* by Giorgio Triulzi
2014-13 *Patents as quality signals? The implications for financing constraints on R&D* by Dirk Czarnitzki, Bronwyn H. Hall and Hanna Hottenrott
2014-14 *Assessment of effectiveness of Chinese aid in competence building and financing development in Sudan* by SamiaSatti Osman Mohamed Nour
2014-15 *Education, training and skill development policies in Arab Gulf countries: Macro-micro overview* by SamiaSatti Osman Mohamed Nour
2014-16 *Structure of labour market and unemployment in Sudan* by SamiaSatti Osman Mohamed Nour
2014-17 *Overview of knowledge transfer in MENA countries - The case of Egypt* by SamiaSatti Osman Mohamed Nour
2014-18 *The impact of ICT in public and private universities in Sudan*by SamiaSatti Osman Mohamed Nour
2014-19 *End-user collaboration for process innovation in services: The role of internal resources* by Mona Ashok, Rajneesh Narula and Andrea Martinez-Noya
2014-20 *Public investment and regional politics: The case of Turkey* by Mehmet Guney Celbis, Denis de Crombrugghe and Joan Muysken
2014-21 *Infrastructure and the international export performance of Turkish regions* by Mehmet Guney Celbis, Peter Nijkamp and Jacques Poot
2014-22 *Discovering and explaining work-family strategies of parents in Luxembourg* by Nevena Zhelyazkova
2014-23 *Parental leave take up and return to work of mothers in Luxembourg: An application of the model of nested dichotomies* by Nevena Zhelyazkova

2014-24 *Millennium Development Goals: Tool or token of global social governance?* by Mueid Al Raee, Elvis Amoateng, Elvis Korku Avenyo, Youssef Beshay, Mira Bierbaum, Charlotte Keijser and Rashmi Sinha

2014-25 *One Europe or several? Causes and consequences of the European stagnation* by Jan Fagerberg and Bart Verspagen

2014-26 *The harmony of programs package: Quasi-experimental evidence on deworming and canteen interventions in rural Senegal* by Théophile Azomahou, Fatoumata Diallo and Wladimir Raymond

2014-27 *Country Terms of Trade 1960-2012: Trends, unit roots, over-differencing, endogeneity, time dummies, and heterogeneity* by Thomas Ziesemer

2014-28 *The structure and comparative advantages of China's scientific research - Quantitative and qualitative perspectives* by Lili Wang

2014-29 *Transition to knowledge-based economy in Saudi Arabia* by Samia Satti Osman Mohamed Nour

2014-30 *Challenges and opportunities for transition to knowledge-based economy in Arab Gulf countries* by Samia Satti Osman Mohamed Nour

2014-31 *Migration of international students and mobilizing skills in the MENA Region* by Samia Satti Osman Mohamed Nour

2014-32 *Beyond product innovation; improving innovation policy support for SMEs in traditional industries* by René Wintjes, David Douglas, Jon Fairburn, Hugo Hollanders and Geoffrey Pugh

2014-33 *The impact of innovation support programmes on SME innovation in traditional manufacturing industries: an evaluation for seven EU regions* by Dragana Radicic, Geoffrey Pugh, Hugo Hollanders and René Wintjes

2014-34 *Beliefs dynamics in communication networks* by Théophile T. Azomahou and Daniel C. Opolot

2014-35 *Stability and strategic diffusion in networks* by Théophile T. Azomahou and Daniel C. Opolot